# Testing cladistics on dialect networks and phyla (gallo-romance and southern italo-romance).

**Antonella GAILLARD-CORVAGLIA**
EA270, Langues Romanes : acquisition, linguistique, didactique, Sorbonne Nouvelle – ParisIII
Antoc75@hotmail.com

**Jean-Léo LEONARD**
UMR 7018
Phonétique et Phonologie.
Sorbonne Nouvelle-ParisIII
75005 PARIS
leonard@idf.ext.jussieu.fr

**Pierre DARLU**
INSERM, U535
Epidémiologie génétique et structure des populations humaines
BP1000 Villejuif, F-94817
darlu@vjf.inserm.fr

## Abstract

This present work deliberately abandons the purpose of capturing the global resemblance between languages and the ambition of giving a rational foundation to probability of changes in linguistics, to focus instead on cladistic approach, which was applied to different dialects and data (gallo-romance, southern italo-romance) through an original coding of philological derivations. Results show good congruence with linguistic classification and provide new insight on how tackle various dialectological problems as borrowings.

## 1 Introduction

In the last decades, theoretical developments in the field of the biological evolution of species and populations have been combined with increasing computer facilities, which are likely to change the practice of phylogeny reconstruction drastically. Attempts to shift such a practice in order to reconstruct the evolution of language have been proposed, since the middle of the 20th century, as evidenced by several publications that display the whole range of methodologies. One of these approaches, called Numerical Taxonomy, consists in estimating some linguistic distances between pairs of languages, from which evolutionary trees or networks are inferred to produce some linguistic classifications. This approach is classically used in dialectometry. (Evrard, 1964; Goebl, 1981, 1987, Scapoli et al., 2005, Ben Hamed, 2005). A more recent approach, based on Bayesian principles, suggests to attach some probabilities to each linguistic change (Gray et al, 2003), looking for the most likely tree, given the model and the observed data. Finally, the last kind of approach, inherited from XIXth century linguists, is the cladistic approach, as formalized by Hennig (1950) and clearly advocated by some linguists, although using various methodologies (Hoenigswald and Wiener, 1987; Wang, 1988; Holden, 2001; Ringe et al., 2002; Rexova et al., 2003; Nakhleh et al., 2005; Ben Hamed et al., 2005).

The present work is focusing on cladistics, abandoning the purpose of capturing the global resemblance between languages and the ambition of giving a rational foundation to probability of linguistic changes, adopting instead a strategy enabling us to integrate linguistic hypotheses before drawing inference on the evolution of linguistic traits and languages, and possibly to refute them. To check the heuristic value of this methodology, we endeavour to apply cladistics to dialectal data from different sources, hoping to bring forward and discuss some arguments on their diversification in space and time. As far as we know, cladistic is more often applied to language families than to dialect areas, so that our research is pioneering the field, raising the controversial question concerning the best representation of dialectal diversity: tree-like and/or networks.

## 2 The data

**2.1 Oïl Dialect[1]**. We began our experiment with the oïl dialects, our starting point being the

[1] Oïl Dialect indicates the branch of the gallo-Romance languages developed in the North of France, south of Belgium (Walloon Area) and inthe Anglo-Normans islands.



*Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, pages 23–30,
Prague, June 2007. 

Linguistic Atlas of France (ALF, Gilliéron and Edmont,1902-1910, reprint: 1968) which has already been extensively exploited by others in a context of global resemblance (Goebl, 1981, 1992). In order to delimit a precise and homogeneous field, the characters observed are limited only to the vocalism of these dialects, mainly stressed and oral vowels, even if a few series of facts from nasal and unstressed vocalism are taken into account. As far as the Oïl data is concerned, the selected localities amount to 45, from East to West, in order to limit our scope in this first attempt (figure 2).

| | |
|---|---|
| ALF | Beau <*bĕllum* (13); Bien<*bĕne* (8); Blé<*blātum* (11); Bœuf <*bŏvem* (5); Cher <*cārum* (9); Eau<*ăquam* (14); Fait <*făctum* (5); Faucille<*fălcīculam* (8); Faux<*fălcem* (15); Feuille<*fŏliam* (11); Fleurs <*flōres* (8); Lit<*lĕctum* (9); Mûr <*matūrum* (14); Mûre <*matūram* (7); Pain<*pănem* (15); Père<*pătrem* (7); Pied<*pĕdem* (13); Poing<*pŭgnum* (10); Pré<*prātum* (16); Puits<*pŭteus* (14);Seigle<*sĕcalem* (21); Tendre<*tĕndere* (7); Toile<*tēlam* (20); |
| ALI | Bocca < *bŭccam* (16); Braccio < *brăchium* (18); Capelli < *capēllos/pĭlos* (19); Dente < *dĕntem* (18); Dito < *dĭgĭtum* (16); Dolce < *dŭlcem* (17); Fegato < *fĭcatum* (15); Forte < *fŏrte* (11); Ginocchio < *genŭculum* (16); Gengiva < *gĕngīvam* (19) Gomito < *cŭbĭtum* (18); Grasso < *grăssum* (10); Grida(lui) <*critāre/allocutāre* (12); Odore < *odōrem* (11); Piede < *pĕdem* (18); Ridere < *rĭdĕre* (16); Sopracciglia <*supercĭlĭum* (19); Sudore < *sudōrem* (17); Vedere < *vĭdēre* (18); Voce < *vŏcem* (16) |

Table 1. Selected words from ALF and ALI Atlases. In parentheses is the number of derivations (states) for each selected world

We selected 23 words from the ALF (Table 1), yielding a variable number of forms or phonetic changes, representing the stressed vocalism of the dialects of Oïl (short/long, high/mid/low vowels in open and close syllabic context).

**2.2 Southern Italo-Romance (SIR).** We then applied the same type of cladistic analysis to the dialects of the dialectal area of Southern Italo-Romance. We made use of the data relating to the consonant system of these dialects, with ALI (Atlante Linguistico dell' Italia, 1995) as a source. In this case, 21 localities were sampled for this analysis, picking up three varieties for each main dialect of these areas (northern, central, and southern: 3 for Campanian, Basilian, Apulian, Calabrian, Sicilian and Salentinian, including also three varieties of Sardinian). The lexical sample amounts to 20 words (Table 1).

## 3 Cladistic analysis

**3.1 Linguistic prolegomena**. From the quoted corpora, diachronic trees were created using the existing bibliography (Chauveau, 1989; Pignon, 1960). But we must reckon and point out that we had a very hard time in trying to make sense out of contradictory or underspecified accounts on chains of phonetic changes available in the literature. We found out – to our bewilderment – that most phonetic changes are quite often telescoped in handbooks of Romance dialectology, monographs, and Ph.D. dissertations, giving only the first and the last stage of phonetic changes: *A > D, instead of *A > *B > *C > D. We therefore had to rely on principles of areologic continuity, as the process of stepping is made hazardous by the vacuum on the successive stages of the sound changes in the literature, in particular in the peripheral varieties of oïl (except in Chauveau's monographs on western Oïl dialects). These principles are the following:

Pr.1. *Principle of areologic continuity*: implies a gradual theory of linguistic change whose stages can be reconstructed on the basis of areal configurations. It entails that stages *B and *C of a *A > D change are available on the maps in current dialects not far from a contiguous centre of gravity. For instance, in western oïl dialects, *e > oi goes through a *e > ei > ai > oi vowel shift whose *ai phase is still to be seen on the ALF maps in the neighbourhood, but it is not akin with the far distant *e > oi change in the East (in Romance lorrain), where the chain *e > ei >oi does not entail an *ai phase.

Pr.2. *Principle of parsimony*: it claims that the vocalic system develops with parsimony the strategies of change; not more than two or three major structural options from which the later evolutions unfold.

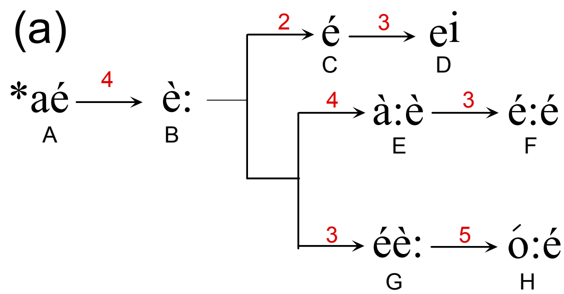
Pr.3. *Principle of unitarianism and naturalness*: dialectal idiosyncrasies should be rare upstream and abundant downstream. In other words, change is strongly constrained typologically closer to the root of the stepping tree, and gets more and more free at the end of the branches. One should be cautious with the intricate complexity of explanations found in monographs and handbooks on idiosyncrasy of changes in local dialects. More simply, one could state that changes are constrained according to UG (Universal Grammar) principles on the first hand, and specified by local, language or dialect-specific parameters on the second hand.

**3.2. Cladistic procedure**. In order to apply cladistic procedure to linguistic data, one has first to find a way to code the trees of philological derivations through a coding procedure which takes into account all the hypotheses assumed by the linguists. In a second step, the field observations have to be coded, and, finally, tree building algorithms are implemented to meet optimal criteria, i.e. parsimony in this context. However, within the framework of this necessarily short paper, we will only discuss tree structure, tackling briefly the feasible reconstruction of ancestral state at nodes, but keeping detailed development for further presentation.

3.2.1 *Character coding*

Figure 1a shows how the relationships between vocalic variations of a given word ("Père", as an example) are coded. Each variant takes the value 0 or 1 depending on its place within the tree derivation. First, a matrix is built (figure 1b), where rows stand for the coding of the variants, whereas columns hold for the transformations from a plesiomorphic variant (the initial diachronic state, or etymon) to an apomorphic one (the terminal state, or synchronic reflexes). For example, the inferred variant, *aé (lettered A), derived from the late latin variant of the A[[2] variable, is coded by the vector [0000000], being the ancestral variant, while its derived reflexes are all coded 1, in the first column. Likewise, the apomorphic variant é:é (lettered F) is coded by the vector [1001100], the first 1 indicates that this variant is derived from the é variant (B), and the fourth and fifth 1 indicate that it is also derived successively from –é– to –à:é– (B−>E) and from –à:é– to – é:é– (E−>F).

(a)




(b)

| | A->B | B->C | C->D | B->E | E->F | B->G | G->H |
|---|---|---|---|---|---|---|---|
| W | 4 | 2 | 3 | 4 | 3 | 3 | 5 |
| A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| B | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| C | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| D | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| E | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| F | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| G | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| H | 1 | 0 | 0 | 0 | 0 | 1 | 1 |

(c)

| From \ to -> | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| A | . | 4 | 6 | 9 | 8 | 11 | 7 | 12 |
| B | i | . | 2 | 5 | 4 | 7 | 3 | 8 |
| C | i | i | . | 3 | i | i | i | i |
| D | i | i | i | . | i | i | i | i |
| E | i | i | i | i | . | 3 | i | i |
| F | i | i | i | i | i | . | i | i |
| G | i | i | i | i | i | i | . | 5 |
| H | i | i | i | i | i | i | i | . |

Figure 1: tree of derivation of the word "Père" (a), its factorized (c) and matrix (d) representations. Each column in (b) corresponds to a change in the tree derivation. The vector W allows a weighting of each shift (example of a 423433 weight-chain, values being expressed above each arrow, and in red colour). The arrows indicate the orientation of changes. Backward changes have an infinity weight. The (c) matrix provides equivalent information, with i holding for infinity weight. (b) and (c) representations are fitted for PHYLIP and PAUP respectively.

Since the transformations can be estimated to be more or less current in term of phonological naturalness, they can be weighed by giving heavy weights for natural or rare transformations and light weights for easy transformations. In this work, character weight was ranged on a scale between 1 and 5 (e.g. w [423433], figure 1).

Lastly, since the transformations are polarized, meaning that we hypothesize the absence of backward changes, we allocate an infinite weight for reversal transformation (i.e. no reversion allowed). This kind of coding is routinely used by phylogeneticists (see PHYLIP or PAUP software). All the derivation trees are available on request.

[2] A[ reads as classical latin low vowel in open syllable as in PA-TREM, MA-REM (noted < [ >, whereas < ] > stands for a closed syllabe as in –AR- : AR-CUM, AR-TEM).

3.2.2. *Data coding and tree reconstruction*

Once the character coding step is performed for all the words investigated (23 different words for the Oïl data (ALF), 20 for SIR (ALI)), each area or dialect is coded according to the previous coding. In the example of the Oïl investigation, data were collected for 45 geographical different areas (figure 2), each of them having its own way to pronounce each of the 23 words of the sample. For example (Table 2), the row « 16Bourg [3]» has the variant C for the first word ("Père") described on the first column. This means that this area as well as the area numbered 45, 59, 65, and 143 share the same derived variant: –é– (labelled C, figure 1), while the rows labelled 108 and 153 share the –è:– reflex variant (B), variants that they inherited either from some common ancestor or because of geographical proximity.

| | |
|---|---|
| 16Bourg | CGLCREJB?NIFFDDEDDDJBBF |
| 45FrCom | CBB?FGEEENCIGIDEDEFJBCF |
| 59LorrRom | CILEMJFBEOFFDCHEDEFEDEF |
| 65FrcomE | CAB?HGE?GNNGKIHEDEFB?BF |
| 108BerNE | BILCGCGEENTFGBECDDDJ?MF |
| 146Champ | CJNHQHLHE?JGGGGEADDF?GC |
| 153Lorr | BILCEHCCEHGJJHHEGEED?BD |

Table 2: Part of the data matrix from the ALF sample. First column is "Père" coding.

Finally, each letter of this data matrix is replaced by its coding (figure 1b), as it has been done in the previous step (character coding). For instance the letter C, column 1 (figure 2) is replaced by the vector [1100000], the letter B by [1000000]. The tree building reconstruction is carried out from this final matrix which sums up all the linguistic hypotheses (tree of philological derivation, polarity of changes, weighting, and geographic variants).

Factorisation are performed with FACTOR software (Felsenstein, 2004), parsimonious trees being obtained with PAUP* (v4.0) (Swofford, 2002), using TBR (tree-bisection-reconnection), random agglomeration option (100), holding 6 best trees at each steps. Tree length, consistency index and retention index are also estimated. The most parsimonious trees are then plotted figures 3 and 4. An example of inferred parallelism is also shown on figure 3. Once clades are well characterized, it becomes possible to count the number of parallelisms that are shared within each clade and those that are shared between clades, giving an estimation of the intensity of borrowing.

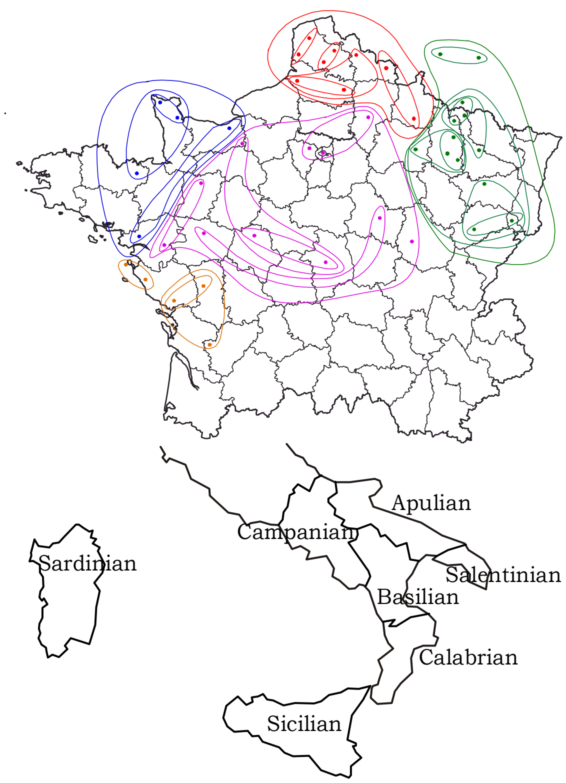



Figure 2: localization of the Oïl and SIR dialect samplings. Contour lines (upper map) correspond to clades from the figure 3.

## 4 Results

4.1. In the O*ïl dialects tree* (figures 2 and 3), the central varieties appear as a clade (C1, from 251Champagne NO to 478Noirmoutier, fuchsia and yellow clades), which gathers the dialects of the Paris basin and those of the mid-west plains, and includes peripheral spots, like Noirmoutier (478) or Saintongeais (518Saintonge). This major node (Center-Western macro-area) makes up a unit of the great mid-west, having the subset Normano-Picardo-Gallo (C2 and C3) as a peripheral compact core. Opposite to this, a very consistent and geographically gradual unit clustering the Franc-Comtois and the Walloon (C4, from 153Lorraine S to 197Wallon O, green clade) varieties, in the Eastern part of the macro-dialect network of Northern Gallo-Romance (i.e., oïl), together with

[3] Number refers to the ALF or ALI areas

the Romance Lorraine[4] dialects. In addition to these great divisions between Central-Western oïl and Peripheric Eastern oïl, which is fairly consistent with current classifications of oïl dialects (Goebl, 1984, 2002), the advantage of this tree lies in the consistency of the inner structures of the major or intermediate clades.

Table 3 gives the estimation of the number of parallelisms and/or borrowings within and between clades. Clearly the number of parallelism observed within each clades turns out to be more intense within than between clades.

| | C1 | C2 | C3 | C4 |
|---|---|---|---|---|
| C1: Fuchsia+Yellow | 3.66 | 1.47 | .084 | 1.19 |
| C2: Red | 1.47 | 4.59 | 1.53 | 1.22 |
| C3: Blue | 0.84 | 1.53 | 9.38 | 1.68 |
| C4: Green | 1.19 | 1.22 | 1.68 | 6.19 |

Table 3: Estimation of the number of parallelisms and/or borrowings within (diagonal) and between clades, standardized by the number of possible exchanges. Clades are defined as figure 3.

4.2. Concerning the *Southern Italo-Romance* (SIR), from Naples to Sicily and Sardinia, the congruence between the cladistic tree (figure 4) and the philological classifications is satisfactory (Goebl, 1984; Grassi et al., 1997), and most of novelties lay in the inner structures of the tree. The phylogram of the SIR shows three major divisions (figure 4): two peripheries, the first one gathering Sardinian Central-Southerner varieties (786 and 748) and the southernmost apulien (818) (red cluster), and the second one (fuchsia) grouping the central-northern apulian (846,828)) as an external branch with the southernmost basilian and central salentino (868 and 917) This last branch is connected to an inner group which separates the branch from Sicilian-Sardinian-Salentino (in blue) from the campano-calabro-basilian (green and yellow). A most interesting detail is the place of 818Apul, a Gallo-Romance francoprovençal dialect settled in two villages (Faeto and Celle) San Vito by the Angevine dynasty in Northern Apulia during the 13th century. This dialect, previously spoken in the Ain and Isère departments in France, got into close, symbiotic contact with Apulian, a dialect of the SIR type. The cladistic procedure grasped accurately its allogenic structure, clustering it in the upper branch, along with Sardinian – also a distinct language as compared to SIR- which should therefore considered as a "foreign languages branch" rather than a peripheric node of the SIR continuum.

## 5 Discussion

The cladistic approach developed here provides a convenient way to integrate and test various hypotheses concerning the linguistic changes. Particularly, the rare or relative absence of backmutation in phonological characters is correctly taken into account by forbidden reverse changes, and complex relationships between states of traits are easily handled, unlike most of the other methods (as network approaches). The parsimony criterion consists to optimise the tree in minimising parallelism. The residual inferred parallelisms could clearly be visualized simply by looking at the places they occur along the tree (as exemplified figure 3). A way to circumvent the parallelism problem, when several parsimonious trees are found, would be using a successive weighting process which looks for parsimonious trees by assigning to each trait a weight inversely proportional to its degree of homoplasy (only parallelism in our case since reversion are not allowed) (Farris 1969). No such a process was necessary with our dataset since only one parsimonious tree was found. However, the robustness of the parsimonious tree remains difficult to evaluate, as long as only few words are integrated in our dataset (only 23 and 20 for Oil and SIR data respectively), particularly to appropriately implement resampling procedures (bootstrap or jackknife).

At this stage of interpretation, one cannot differentiate between parallel development and borrowing, unless some *a priori* are introduced to do so. In our data set, parallelisms are frequent (leading to a weak CI) although our two parsimonious trees are unique and well resolved (actually, there is no simple relation between CI and tree resolution) preventing us using various

[4] As opposed to German Lorrain dialects (Lower-German type) spoken around Metz, whereas Romance Lorrain oïl dialects are or were spoken around Nancy and in the Vosges hills.

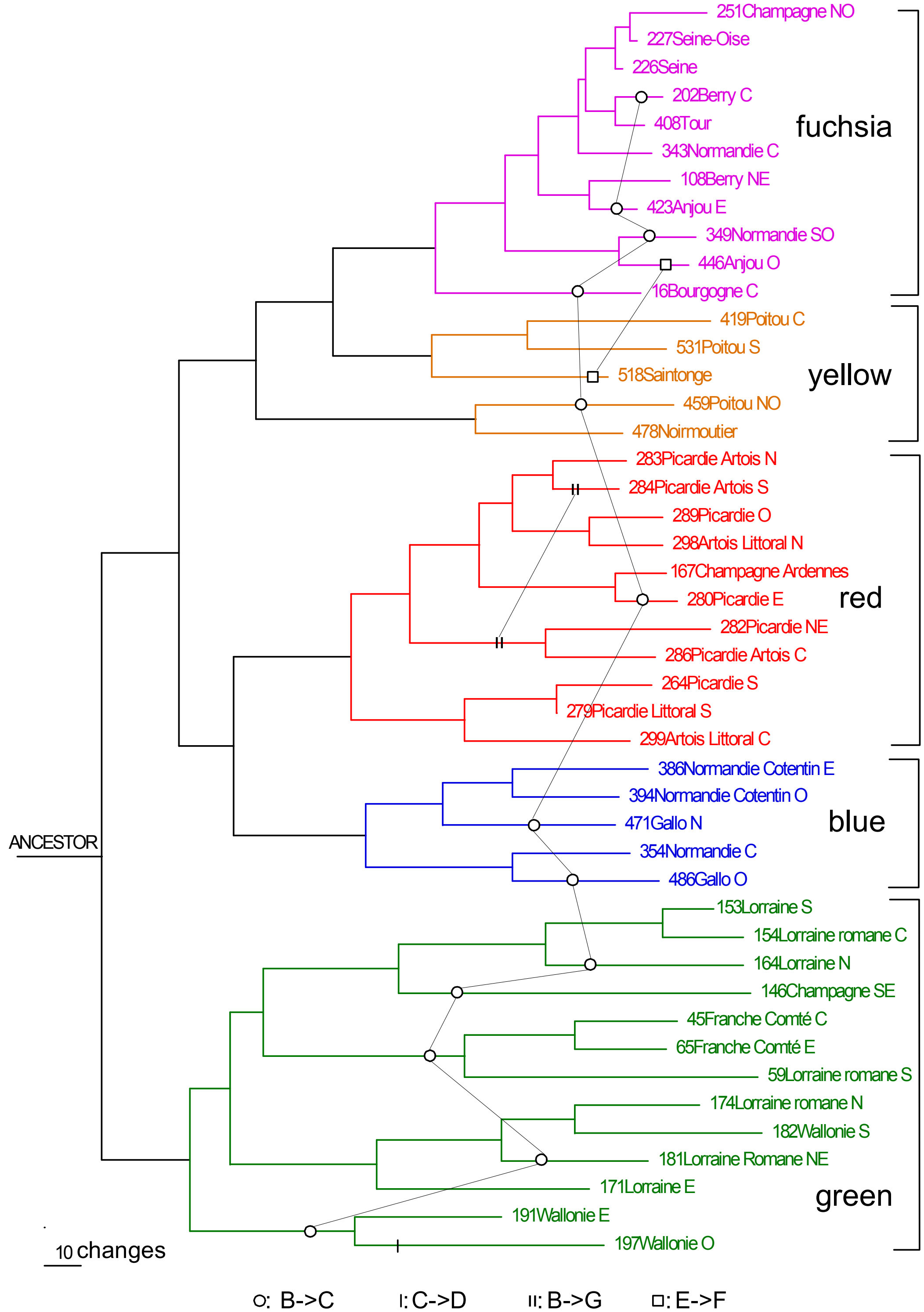



Figure 3. Oïl dialect parsimonious tree (tree length= 2558; Consistency Index (CI)=0.29; CI excluding uninformative characters = 0.22. Retention index (RI) = 0.74; Rescaled consistency index (RC) = 0.21). Branch lengths are proportional to the number of changes. Dialect numbers are labelled as in ALF. Parallel changes for “Père” are localized on the branches (see also figure 1)


Figure 4 : SIR's parsimonious tree (Parsimonious tree: length = 1529, Consistency index (CI) = 0.59 ; CI excluding uninformative characters =0.40 ; Retention index (RI) = 0.53 ; Rescaled consistency index (RC) = 0.32) (Apu: Apulian;Si:Sicilian; Sar:Sardinian; Cal:Calabrian; Cam:Campanian; Bas: Basilian. N: North; C: Centre; S:South. (see also figure 2)

network approaches[5] (median network or median joining based on characters, definitively excluding neighbornet method which is based on global resemblance and is in any way cladistic). These approaches are not able, as far as we know, to handle large amount of polarized changes and complex weighted multistate relationships. On the other hand, our strategy turns out to be quite different from the one proposed by Nakhelh et al (2006) which first apply compatibility method to select the best traits allowing to retain few trees considered as "almost perfect phylogenies" (missing the phylogenetic information brought by the other traits), and then to parsimoniously handle the remaining traits as possible edges representing borrowing, (but not giving the possibility of modifying the tree structure accordingly). An additional advantage (only lightly evoked in this paper) of the cladistic approach is to allow inferring changes of the traits along the tree, suggesting some linguistic scenarios, as correlated changes, borrowings …

[5] See SPLITTREE and NETWORK packages in ref.

## 6 Conclusions

We shall conclude this pioneering cladistic survey of phyla and dialect networks pointing out at three main assets of our data processing : i) unlike most of current and past research in taxonomy applied to linguistic data, we tried to do much more than merely computing distance and similarity between lists of lexical cognates with a binary procedure: we processed data according to geolinguistic analysis, using area linguistic procedures and phonological markedness theory in endowing weight to reflexes, ii) our results are mainly congruent and consistent with current classification, but intricate patterns in the inner structures of cladistic nodes also challenging these classifications, iii) In spite of the small number of words presently studied here, but thanks to accurate data and proper sampling from the ALF and ALI database, it turns out that, by applying cladistics, for long advocated by linguists, one can obtain consistent, reliable (and possibly refutable) results. This is not always the case in the processing of fuzzy data and mere lists of words.

Linguists and cladisticians should therefore be cautious about word-lists, and should as well rely on linguistic atlases, which provide the widest array of sampling, and high quality data gathered through fieldwork by highly trained professionals. In other words, to put it straightforwardly, well managed empiricism is *a sine qua non* condition for reliable results in quantitative linguistics, especially as far as cladistics is concerned, due to the powerfulness of the procedure.

## 7 References


ALI, *Atlante Linguistico Italiano*. 1995-1996. Istituto Poligrafico dello Stato, vol. I-II.

ALF, Gilliéron J, Edmont E. 1902-1910. *Atlas linguistique de France*, Paris, 10 vol. (re-edition: Boulogne 1968)

Ben Hamed M., Darlu P., Vallée N. 2005. On Cladistic reconstruction of linguistic trees through vocalic data. *J. of Quantitative Linguistics*,12(1) :79-109

Ben Hamed M., 2005. Neighbour-nets portray the Chinese dialect continuum and the linguistic legacy of China's demic history. *Proceedings of the Royal Society of London B* , 272:1015–1022.

Chauveau J.-P. 1989. *Evolutions phonétiques en gallo,* Paris : CNRS (coll. Sciences du Langage) ,293 p.

Evrard E. 1964. Etude statistique sur les affinités de cinquante-huit dialectes Bantous." in: *Statistique et Analyse linguistique*, Colloque de Strasbourg 20-22 avril 1964. Presses Universitaires de France, 1966

Farris J.S. 1969. A successive approximations approach to character weighting. *Sys Zool* 18:374-85

Felsenstein J. 2004. PHYLIP (Phylogeny Inference Package) version 3.6b. Department of Genome Sciences, University of Washington, Seattle.

Goebl H. 1981. Eléments d'analyse dialéctometrique (avec application à l'AIS). *Revue de Linguistique Romane*, 45, 349-420

Goebl H 1984. Dialektometrische Studien. Anhand italoromanischer, rätoromanischer und gallo-romanischer Sprachmaterialien aus AIS und ALF, Tübingen (Niemeyer), 3 vol. 254S., 379S., 289S.

Goebl H. 1987. Points chauds de 'analyse dialectométrique: pondération et visualization. *Revue de Linguistique Romane*, 51 :63-118.

Goebl H 1992. Problèmes et méthodes de la dialectométrie actuelle. *IKER 7*. Euskaltzaindia. Real Academia de la lengua Vasca, Bilbao, pp429-475.

Goebl H 2002. Analyse dialectométrique des structures de profondeur de l'ALF. *Revue de linguistique romane*. 261-262 : 5-64

Grassi C, Sobrero A, Telmon T, 1997. Fondamenti di dialettologia italiana, Editori Laterza, Bari.

Gray R.D, Atkinson Q.D. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* |( 426): 435-437

Hennig W 1950. *Grundzüge einer Theorie der Phylo-genetischen Systematik*. Deutscher Zentralverlag (Berlin)

Hoenigswald M. H. and Wiener F L 1987. *Biological Metaphor and Cladistic Classification: An Interdisciplinary Perspective.* University of Pennsylvania Press, Philadelphia.

Holden C.J. 2001. Bantu language trees reflect the spread of farming across sub-saharian Africa: a maximum parsimony analysis. *Proceeding of the Royal Society (London)* 269:793-799.

Nakhleh L.; Ring D., Warnow T. 2005. Perfect phylogenetic networks: a new methodology for reconstructing the evolutionary history of natural languages. *Language* 11(2):382-420.

NETWORK v4.2.0.1, 2007. www.fluxusengineering.com/network_terms.html

Pignon E. 1960. L'évolution phonétique des parlers du Poitou (Vienne et Deux-Sèvres). Editions d'Artrey, Paris.

Rexova K., Frynta D., Zrzavy J. 2003. Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics* (19):120-127.

Ringe D, Warnow T., Taylor A., 2002. Indo-European and computational cladistics. *Transaction of the philological society.* 100(1):59-129.

Scapoli C., Goebl H., Sobota S., Mamolini E., Rodriguez-Larralde A., Barraï I. 2005. Surnames and dialects in France: piopulation structure and cultural evolution. *J Theor Biol.* 237(1):75-86.

SPLITTREEv4beta. 2005. www.splittree.org

Swofford, D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4.0. Sinauer Associates, Sunderland, Massachusetts.

Wang W.S.Y. 1987. Representing languages relationships. In: Hoenigswald, M. H. and Wiener, F. L. (1987) " *Biological Metaphor and Cladistic Classification : An Interdisciplinary Perspective*" University of Pennsylvania Press, Philadelphia